Malaria Genetics and Genomics

Anopheles gambiae (African malaria mosquito) is the primary vector that transmits *Plasmodium falciparum*, the causative agent of human malaria. NCBI provides direct access to the genomic sequence data of *A. gambiae*, *P. falciparum* and other malaria-associated genomic resources described below and featured at **www.ncbi.nlm.nih.gov/projects/Malaria/**.
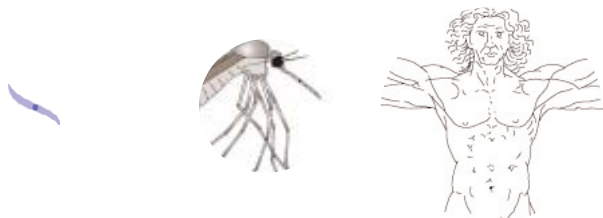
Content of Entrez Databases

Nucleotide: Genomic DNA and cDNA sequences from GenBank and NCBI Reference Sequences.

Protein: Protein sequences from PIR, PRF, SWISS-PROT, PDB, conceptual translations of GenBank sequences and NCBI Reference Sequences.

Structure: Experimentally determined 3D structures from Protein Data Bank (PDB).

Genomes: Genomic sequences for *P. falciparum*, *A. gambiae, H. sapiens*, and many other organisms.

SNP: DNA polymorphisms (SNPs, insertions, deletions, microsatellites, etc.).

UniGene: Gene-oriented clusters of expressed sequence tags (ESTs) and mRNA sequences.

UniSTS: A database of Sequence Tagged Sites.

PopSet: Global alignments of DNA sequences from population studies.

Reference Sequences www.ncbi.nlm.nih.gov/RefSeq

NCBI Reference Sequences (RefSeqs) comprise a curated, non-redundant set of sequences derived from the primary sequence data in GenBank. RefSeqs include curated mRNA (NM_######) and protein (NP_######) sequences, as well as those of whole chromosomes (NC_######), among others. For *P. falciparum*, NC records in Entrez Genomes represent the finished chromosomes and were assembled from GenBank sequences submitted by the Malaria Genome Project (MGP) Consortium. The many NP records were derived from the coding sequence annotations provided on the relevant GenBank sequences. For *A. gambiae* the mitochondrial genome is available as an NC record, while the nuclear genome is available in the Map Viewer and as a set of NCBI "NT_" contigs.

dbEST www.ncbi.nlm.nih.gov/dbEST/index.html

A bulk sequence division of GenBank containing ESTs that represent short (200-500 bp), single-read sequence data from expressed genes in a cell, tissue, or organ under specific cellular conditions.

Trace Archive www.ncbi.nlm.nih.gov/Traces/trace.cgi?

A repository of raw sequence data from sequencing centers.

Indexing of *A. gambiae* and *P. falciparum* Genomic Sequences

A. gambiae: The genome sequencing project has accession **AAAB00000000**. The current assembly shown in the Map Viewer is Version 2.

P. falciparum: The individual chromosomes are each represented by one NC Reference Sequence.

Data Access Using Entrez: Entrez Queries

From any Entrez database:

Anopheles gambiae[organism]
Plasmodium falciparum[organism]
Human[organism]

For Reference Sequences (RefSeqs) in Entrez Nucleotide or Protein:

Plasmodium falciparum[organism] AND srcdb refseq [properties]

For ESTs in Entrez Nucleotide:

Anopheles gambiae[organism] AND gbdiv est[properties]

Data Access Using FTP

ftp.ncbi.nih.gov/pub/Malaria
ftp.ncbi.nih.gov/genbank/genomes/Anopheles_gambiae
ftp.ncbi.nih.gov/genbank/genomes/P_falciparum
ftp.ncbi.nih.gov/genomes/H_sapiens

Data Records in Selected Databases Held at NCBI as of October 8, 2003			
	A. gambiae	*P. falciparum*	*H. sapiens*
Entrez			
Nucleotide	342,433	28,889	7,120,334
Protein	33,105	16,020	202,690
Structure	1	38	4,530
Genomes	6	15	25
SNP	976,063	172	5,875,084
UniGene	13,621	0	123,995
UniSTS	0	893	174,410
PopSet	25	25	320
RefSeq			
Nucleotide	17,449	15	43,650
Protein	16,106	5,270	41,528
dbEST	130,731	21,305	5,426,016
Trace Archive	5,081,012	224,023	26,715,332
Genomic Assembly	Yes	Yes	Yes

Map Viewer

P. falciparum

The NCBI Map Viewer for *P. falciparum* displays genetic information for all 14 chromosomes using the following five maps:

Contig: displays the NC Reference Sequence record for the chromosome.

Component: displays the tiling path of GenBank sequences used to assemble the NC record.

Gene: displays known or putative genes based on alignments of the NP records to the NC records using blastx.

STS: placement of STSs from UniSTS onto the contig sequence map using Electronic-PCR (e-PCR).

NIAID Genetic: The malaria genetic linkage map from (Su et al. 1999), using markers and map information available from the NCBI Malaria Genetics and Genomics site. Microsatellite markers are position-based on their recombination rates.

Py Proteins: displays *P. yoelii* proteins aligned to *P. falciparum* chromosomes.

www.ncbi.nlm.nih.gov/mapview/map_search.cgi?chr=pfalciparum.inf

A. gambiae

The NCBI Map Viewer for *A. gambiae* displays the following cytogenetic and sequence maps:

Satellites: sequence map of satellite positions.

Satellites on Bands: cytogenetic map of satellite positions.

Ideogram: ideogram of the chromosome with cytogenetic coordinates.

Component: sequence map of intermediate genomic assemblies of WGS sequence data.

Component on Bands: cytogenetic map of component positions.

Contig: sequence map of contig positions.

Gene: sequence map of annotated *A. gambiae* genes with links to protein sequences.

Transcript: sequence map of *A. gambiae* predicted gene models aligned to the genome assembly.

www.ncbi.nlm.nih.gov/mapview/map_search.cgi?chr=agambiae.inf

BLAST

A. gambiae

NCBI provides three ways to perform BLAST searches against *A. gambiae* genomic data:

Genomic BLAST: provides BLAST searches against *A. gambiae*, *D. Melanogaster*, or both genomic scaffolds or proteins. The available versions of BLAST are blastp, blastn, blastx. tblastn, and MegaBLAST.

www.ncbi.nlm.nih.gov/BLAST/Genomes/FlyBlast.html

Trace Archive BLAST: allows BLAST searches against the raw trace data for *A. gambiae* using MegaBLAST and is optimized for searching with *A. gambiae* queries.

www.ncbi.nlm.nih.gov/blast/mmtrace.html

Cross-species Trace Archive BLAST: allows BLAST searches against the raw trace data for *A. gambiae* using discontiguous MegaBLAST and is optimized for searching with queries from other species.

www.ncbi.nlm.nih.gov/blast/tracemb.html

P. falciparum

NCBI provides three ways to perform BLAST searches against *P. falciparum* genomic data, as well as genomic data for other *Plasmodium* species:

Genomic BLAST: provides access to both DNA and protein databases for *P. falciparum*, *P. yoelii*, or both organisms. The available versions of BLAST are blastp, blastn, blastx, tblastn, and Megablast.

www.ncbi.nlm.nih.gov/BLAST/Genomes/plasmodium.html

GenBank BLAST: enables searches against selected GenBank divisions (EST, GSS, STS, and HTGS), along with the standard nr database, or a combination of all of the above. Searches can be restricted to *P. falciparum*, all *Plasmodium*, or all toxoplasma. Only nucleotide databases are available here, so the BLAST versions provided are blastn, tblastn, and tblastx.

www.ncbi.nlm.nih.gov/projects/Malaria/plasmodiumbl.html

Custom BLAST: provides BLAST access to unfinished sequences provided by the MGP Consortium but not necessarily in GenBank. Databases for individual chromosomes are available for *P. falciparum*, while broader databases are available for *P. berghei*, *P. vivax*, *P. chabaudi*, *P. knowlesi*, and *P. yoelii*.

www.ncbi.nlm.nih.gov/projects/Malaria/plasmodiumblcus.html

Other Useful NCBI Resources

Rodent Malaria Resources: This page provides data and information relevant to the four species of *Plasmodium* that cause malaria in rodents: *P. berghei*, *P. chabaudi*, *P. yoelii,* and *P. vinckei*.

www.ncbi.nlm.nih.gov/projects/Malaria/Rodent/index.html

PubMed: Access to the Biomedical Literature. PubMed, a service of the National Library of Medicine, provides free, on-line access to over twelve million abstracts indexed in MEDLINE. Key terms in each abstract are hyperlinked to Entrez Books for additional information content. When available, free electronic versions of published reports can be downloaded from PubMed Central. LinkOut provides links to related journal entries and additional resources within and outside of NCBI.

www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed